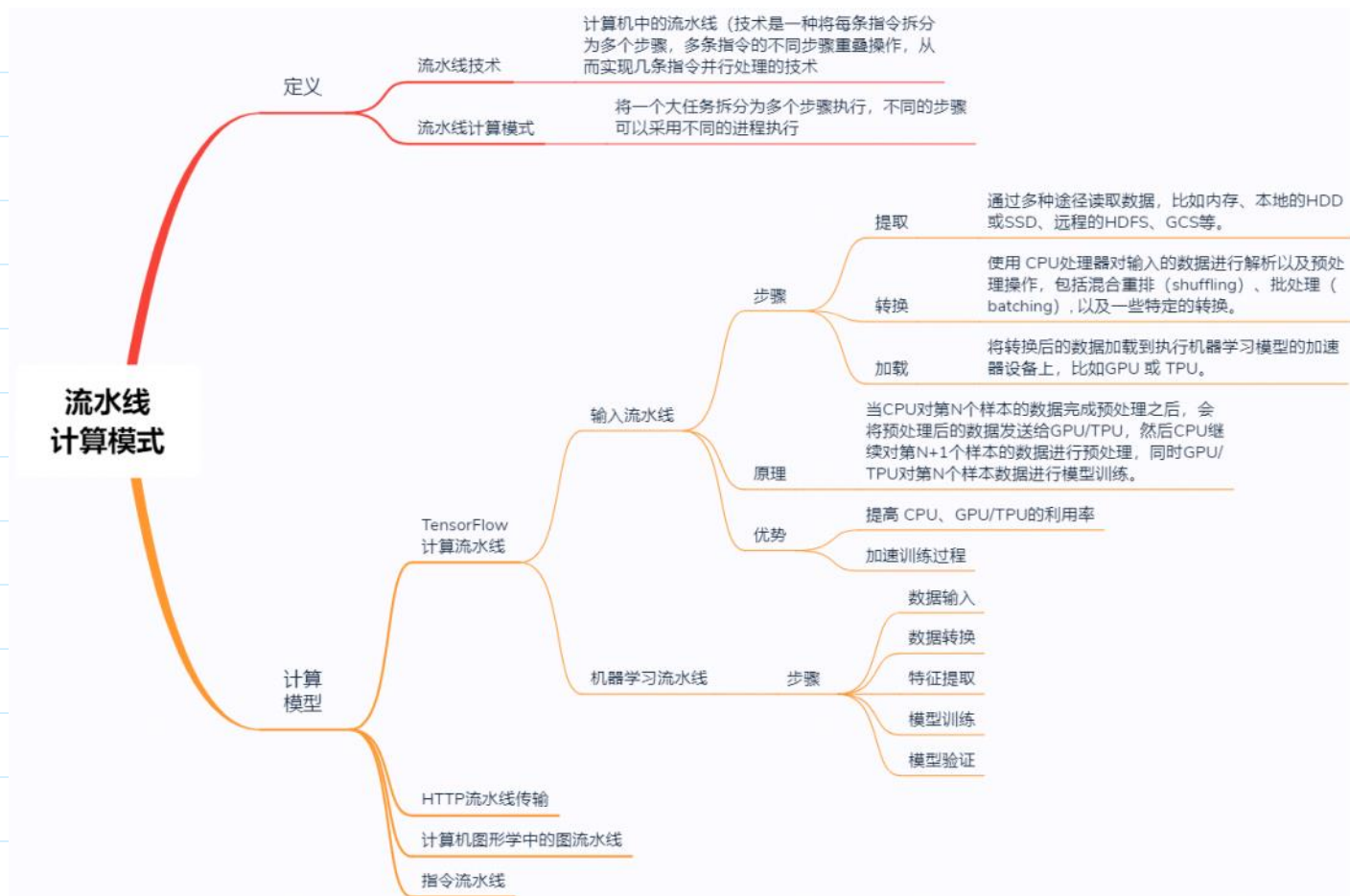


## 3.4 流水线

2022年3月7日 17:53



流水线思想。  
一整件事拆成  $n$  个子任务，子任务重叠运行。

流水线计算模式原理。

TensorFlow 运用了流水线模式对输入数据进行预处理，因此称为输入流水线（TensorFlow Training Input Pipelines）。其数据输入流水线主要包含 3 个步骤：

- **提取 (Extract)**。通过多种途径读取数据，比如内存、本地的 HDD 或 SSD、远程的 HDFS、GCS 等。数据的种类也有很多，比如图像数据、文本数据、视频数据等。
- **转换 (Transform)**。使用 CPU 处理器对输入的数据进行解析以及预处理操作，包括混合重排 (shuffling)、批处理 (batching)，以及一些特定的转换。比如图像解压缩和扩充、文本矢量化、视频时序采样等。
- **加载 (Load)**。将转换后的数据加载到执行机器学习模型的加速器设备上，比如 GPU 或 TPU。